

Multilevel Monte Carlo for the Feynman-Kac Formula for the Laplace Equation

Stefan Pauli^{*1,2}, Robert Gantner¹, Peter Arbenz¹, and Andreas
Adelmann³

¹ETH Zurich, Computer Science Department, 8092 Zurich,
Switzerland

²ETH Zurich, Seminar for Applied Mathematics, 8092 Zurich,
Switzerland

³PSI, 5232 Villigen, Switzerland

January 17, 2014

Abstract

Since its formulation in the late 1940s, the Feynman-Kac formula has proven to be an effective tool for both theoretical reformulations and practical simulations of differential equations. The link it establishes between such equations and stochastic processes can be exploited to develop Monte Carlo sampling methods that are effective, especially in high dimensions. There exist many techniques of improving standard Monte Carlo sampling methods, a relatively new development being the so-called Multilevel Monte Carlo method. This paper investigates the applicability of multilevel ideas to the stochastic representation of partial differential equations by the Feynman-Kac formula, using the Walk on Sphere algorithm to generate the required random paths. We focus on the Laplace equation, the simplest elliptic PDE, while mentioning some extension possibilities.

1 Introduction

The Monte Carlo (MC) error normally converge like $1/\sqrt{M}$, where M is the number of samples. If all samples were equally expensive the error versus work convergence rate would be $1/\sqrt{W}$, where W is the total work. Note

^{*}The work of this author has been funded by the ETH interdisciplinary research grant CH1-03 10-1.

that for a large class of problems the accuracy of the MC method is not only determined by the number of samples used, but also by the discretization error accepted in the computation of each sample. A small discretization error, using e.g. the Walk on Spheres (WoS) algorithm [14], comes with larger computation costs per sample. Hence the error versus work convergence rate falls short of $1/\sqrt{W}$. If applicable, Multilevel Monte Carlo (MLMC) methods may reach this optimal $1/\sqrt{W}$ convergence rate. This motivated us to evaluate the Feynman-Kac formula with MLMC methods instead of MC methods.

In this paper we present a procedure to evaluate a Feynman-Kac formula with MLMC using the Walk on Spheres method. As a model problem we use the Laplace equation, which is solved in high dimensions for instance for option pricing [11, Chap. 8] or in particle accelerator simulations [1]. We prove that the error versus work converges with the optimal $1/\sqrt{W}$ convergence rate, and compute the suboptimal convergence when using MC methods. The MLMC method outperforms the MC method only by a $\log(W)$ term. Our MLMC simulations, executed with an MPI parallel implementation, where up to twice as fast compared to the standard MC implementation.

First, in Section 2 we introduce the Feynman-Kac formula, using the WoS algorithm, and compute the error versus work convergence rate when using MC. Then, we derive a method to use MLMC in this setting, and evaluate its error versus work convergence rate in Section 3. In Section 4 we present numerical results, which quantify the advantage of using MLMC. Finally, we draw our conclusions in Section 5.

2 Standard Walk on Spheres

2.1 Brownian motion

A Brownian motion [6] denoted by X_t , started at a point in a certain connected domain D , has several characteristic quantities. These include the first exit time and the first exit point, two important concepts in the application of the Feynman-Kac formula.

Definition 2.1 (First exit time). *The time at which a realization of a Brownian motion X_t , started at some point $x \in D \subset \mathbb{R}^d$, first hits the domain boundary ∂D is called the first exit time $\tau := \inf\{t > 0 : X_t \in \partial D\}$.*

Definition 2.2 (First exit point). *The point at which a realization of a Brownian motion X_t , started at some point $x \in D \subset \mathbb{R}^d$, first hits the domain boundary ∂D is called the first exit point X_τ .*

Of great importance for the derivation of the Walk on Spheres algorithm is the *distribution* of first exit points of a Brownian motion in a ball $B_x(R)$ of radius R centered at the point x . If the starting point of the Brownian

motion is the center of the ball, this distribution is uniform on its spherical surface, independent of the dimension [18, Theorem 3].

In evaluating functionals of stochastic processes, one is often interested in their expectations and variances. Throughout this paper, the L^2 -norm over the sample space Ω is defined as the expectation $\|X\|_{L^2(\Omega)} := \mathbb{E}[|X|^2]^{\frac{1}{2}}$.

2.2 Feynman-Kac formula for the Laplace equation

The Feynman-Kac formula, developed by Richard Feynman and Mark Kac [12], gives a probabilistic representation of the solution to certain PDEs at a single fixed evaluation point x . The Feynman-Kac formula allows us to write a very general elliptic PDE

$$\begin{aligned} Lu &= \frac{1}{2}\Delta u + \sum_{i=1}^d b_i(x)\partial_i u + c(x)u = -g(x), & \text{in } D \subset \mathbb{R}^d, \\ u(x) &= f(x), & \text{on } \partial D, \end{aligned}$$

along with the assumption that b , c and g are smooth and satisfy a Lipschitz growth condition [5] as an expectation over a stochastic process X_t beginning at the point x . This process fulfills the stochastic differential equation $dX_t = b(x)dt + dW$ where dW is a Brownian increment, and is stopped as soon as it hits the boundary. Denoting the first exit time by τ and the condition that $X_0 = x$ by the subscript x in \mathbb{E}_x , the formula can be written as

$$u(x) = \mathbb{E}_x \left[f(X_\tau) \exp \left(\int_0^\tau c(X_s) ds \right) + \int_0^\tau g(X_t) \exp \left(\int_0^t c(X_s) ds \right) dt \right]. \quad (1)$$

In this paper we are only interested in the Laplace equation $\Delta u = 0$ in d dimensions on a domain $D \subset \mathbb{R}^d$ with boundary values $u(x) = f(x)$ on ∂D . This simplifies the development of an MLMC method, which can be used as a basis for more general equations. In this simple case the Feynman-Kac representation is given as an expectation over the exit point X_τ of a Brownian motion started at the point x . Denoting the condition $X_0 = x$ by the subscript x in \mathbb{E}_x , the solution is written as [5]

$$u(x) = \mathbb{E}_x[f(X_\tau)] =: \mathbb{E}_x[Y]. \quad (2)$$

2.3 Walk on Spheres (WoS)

The Walk on Spheres (WoS) algorithm [14] can be viewed as an alternative to a conventional detailed simulation of the drift-free Brownian motion inside the domain D . Started at the center of a sphere the exit point distribution is known. This allows to simulate the Brownian motion using discrete jumps of a given size. Starting at $\tau = 0$ with $X_0 = x$ the algorithm measures the distance from the current position X_t to ∂D and jumps this distance

in a uniformly random direction to the next position X_{t+1} . The algorithm terminates when it is ε close to the boundary ∂D at the point X_N , where N denotes the number of WoS steps needed. The first exit point is approximated by the point $\bar{X}_N \in \partial D$ that is closest to X_N . This WoS algorithm may be used to solve the Laplace equation [14]. We recall that $\mathbb{E}[Y] = u(x)$ when Y is computed with a realization of a processes starting at x . $Y_\varepsilon := f(\bar{X}_N)$, is the estimator based on an WoS processes with the discretization parameter ε . The discretization error presented in [13] is given as

$$e_{\text{discr}} = \|\mathbb{E}[Y] - \mathbb{E}[Y_\varepsilon]\|_{L^2(\Omega)} \leq \mathcal{O}(\varepsilon). \quad (3)$$

In [3], a worst-case upper bound of $\mathcal{O}(\varepsilon^{4/d-2})$ for three or more dimensional problems ($d \geq 3$) is derived for the expected WoS path length $\mathbb{E}[N]$. For domains fulfilling certain regularity conditions tighter upper bounds of $\mathbb{E}[N] = \mathcal{O}(\log^p(\varepsilon^{-1}))$, $d \geq 2$ are proven for $p = 1$ or $p = 2$ depending on the domain. These estimates are relevant for bounding the expected work per sample.

2.4 Monte Carlo

2.4.1 Statistical error

Given M realizations $\{Y_\varepsilon^i\}_{i=1}^M$ of the random variable Y_ε obtained from independent paths of the WoS process, the value for $\mathbb{E}[Y_\varepsilon]$ can be approximated by the estimator $E_M[Y_\varepsilon] = \frac{1}{M} \sum_{i=1}^M Y_\varepsilon^i$. The error of this Monte Carlo estimator can be written in terms of the random variable Y_ε as

$$\begin{aligned} e_{\text{stat}} &= \|\mathbb{E}[Y_\varepsilon] - E_M[Y_\varepsilon]\|_{L^2(\Omega)} = \text{Var} \left[\frac{1}{M} \sum_{i=1}^M Y_\varepsilon^i \right]^{\frac{1}{2}} \\ &= \frac{1}{\sqrt{M}} \sqrt{\text{Var}[Y_\varepsilon]} \leq \frac{1}{\sqrt{M}} \|Y_\varepsilon\|_{L^2(\Omega)}, \end{aligned} \quad (4)$$

where we use the fact that the samples Y_ε^i are independent and identically distributed (i.i.d.) realizations of Y_ε .

2.4.2 Total error

The total error of an estimation $E_M[Y_\varepsilon]$ of $\mathbb{E}[Y]$ by the Walk on Spheres algorithm can be written using the triangle inequality of the $L^2(\Omega)$ norm as

$$\|\mathbb{E}[Y] - E_M[Y_\varepsilon]\|_{L^2(\Omega)} \leq \|\mathbb{E}[Y] - \mathbb{E}[Y_\varepsilon]\|_{L^2(\Omega)} + \|\mathbb{E}[Y_\varepsilon] - E_M[Y_\varepsilon]\|_{L^2(\Omega)},$$

resulting in a total error of

$$e_{\text{tot}} := \|\mathbb{E}[Y] - E_M[Y_\varepsilon]\|_{L^2(\Omega)} \leq \mathcal{O}(\varepsilon) + \sqrt{\frac{\text{Var}[Y_\varepsilon]}{M}}. \quad (5)$$

The two terms at the right follow by (3) and (4), respectively.

2.4.3 Error equilibration

For a fixed prescribed error, one way of choosing the sample size is by equilibrating the statistical and discretization errors in (5), $\mathcal{O}(\varepsilon) = \sqrt{\text{Var}[Y_\varepsilon]}/M$. This yields the relationship

$$M = \mathcal{O}(\varepsilon^{-2}), \quad (6)$$

giving a total error behavior of $\mathcal{O}(\varepsilon)$.

2.5 Error vs. expected work

The expected total work $\mathbb{E}[W]$ of a WoS simulation is the number of paths times the expected length of a path:

$$\mathbb{E}[W] = M \cdot \mathbb{E}[N]. \quad (7)$$

Each bound for the expected path length $\mathbb{E}[N]$ shown in section 2.3 is multiplied with M derived in (6), and the resulting expected work is solved for ε . Equation (8a) is valid for $d \geq 3$, equation (8b) for well-behaved domains $d \geq 2$. To derive (8b) we must make use of the Lambert W -function, $\mathcal{W}_{\text{lam}}(\cdot)$, defined as the inverse of the map $w \mapsto w \exp(w)$. It can be approximated with the truncated expansion $\mathcal{W}_{\text{lam}}(x) \approx \log(x) - \log \log(x)$ [4]. This yields the following relationships:

$$\mathbb{E}[W] = \mathcal{O}(\varepsilon^{4/d-4}) \quad e_{\text{tot}} = \mathcal{O}(\varepsilon) = \mathcal{O}(\mathbb{E}[W]^{\frac{1}{4} \frac{d}{1-d}}) \quad (8a)$$

$$\mathbb{E}[W] = \mathcal{O}(\varepsilon^{-2} \log^2(\varepsilon^{-1})) \quad e_{\text{tot}} = \mathcal{O}(\varepsilon) = \mathcal{O}(\mathbb{E}[W]^{-\frac{1}{2}} \log(\mathbb{E}[W])) \quad (8b)$$

The total error is linear in ε by the choice in (6). In both cases the WoS algorithm performs worse than the optimum possible in a Monte Carlo setting, namely $W^{-\frac{1}{2}}$. This motivates the formulation of a multilevel version of the WoS algorithm, with the hopes of achieving the optimal convergence rate.

3 Multilevel Walk on Spheres (MLWoS)

In this section, a multilevel version of the WoS algorithm is formulated and its error behavior analyzed. The main idea is to execute the WoS algorithm on different “discretization levels”, meaning for different values of the discretization parameter ε . The subscript $\ell \in \{0, \dots, L\}$ is used to denote a certain discretization level, where $\ell = 0$ is the coarsest discretization level, corresponding to a WoS simulation with ε_0 , and $\ell = L$ is the finest discretization level.

3.1 Multilevel formulation

On each discretization level $\ell = 0, \dots, L$, we define the discretization parameter $\varepsilon_\ell := \eta^{-\ell} \varepsilon_0$, where $\eta > 1$, i.e. the width of the stopping region is divided by η between successive levels.

In the following definition, a single WoS process with ε_ℓ is used to define the multilevel process in order to incorporate the fact that stopping points on higher levels are continuations of previously stopped processes.

Definition 3.1 (Multilevel Walk on Spheres process). *Given a domain $D \subset \mathbb{R}^d$, a point $x \in D$, a discretization parameter ε_0 and an $\eta > 1$. Consider a WoS process $\{X_i\}_{i=0}^{N_\ell}$ started at x with $\varepsilon_\ell = \eta^{-\ell} \varepsilon_0$ and the pair $(X_{N_{\ell-1}}, X_{N_\ell})$ obtained by $N_{\ell-1} = \min\{N \in \mathbb{N} : d_{\partial D}(X_N) < \varepsilon_{\ell-1}\}$, where $d_{\partial D}(X_N) = \min_{x' \in \partial D} |X_N - x'|$ is the distance to the boundary ∂D . The Multilevel Walk on Spheres process (MLWoS) on the level ℓ is the set of all such pairs.*

In the context of multilevel Monte Carlo, we have multiple estimators Y_{ε_ℓ} , one for each discretization level with discretization parameter ε_ℓ . These are obtained in the same way as for the non-multilevel case, e.g. $Y_{\varepsilon_\ell} = f(\bar{X}_{N_\ell})$. The expectation of the estimator Y_{ε_L} is written in a multilevel form as

$$\mathbb{E}[Y_{\varepsilon_L}] = \mathbb{E}[Y_{\varepsilon_0}] + \sum_{\ell=1}^L \mathbb{E}[Y_{\varepsilon_\ell} - Y_{\varepsilon_{\ell-1}}]. \quad (9)$$

Replacing the expectation with the average over M_ℓ realizations on each discretization level ℓ , we get the MLMC estimate

$$E[Y_{\varepsilon_L}] = E_{M_0}[Y_{\varepsilon_0}] + \sum_{\ell=1}^L E_{M_\ell}[Y_{\varepsilon_\ell} - Y_{\varepsilon_{\ell-1}}]. \quad (10)$$

The “sample” on level $\ell > 0$ is now $(Y_{\varepsilon_\ell} - Y_{\varepsilon_{\ell-1}})$, for which it is assumed that the two values on discretization levels ℓ and $\ell - 1$ come from the same WoS path. The implication is that they should be simulated with the MLWoS process from 3.1. The remaining estimator on discretization level 0 is computed with an ordinary WoS simulation.

Note that the individual Y_{ε_ℓ} are estimators on the *discretization level* ℓ , whereas $Y_{\varepsilon_\ell} - Y_{\varepsilon_{\ell-1}}$ are referred to as estimators on *level* ℓ .

3.2 Multilevel error bounds

The multilevel Monte Carlo error is defined as the difference between the expectation of the exact estimator $\mathbb{E}[Y]$ and the MLMC approximation involving all levels, $E[Y_{\varepsilon_L}]$, and is given by [2]

$$\begin{aligned} \|\mathbb{E}[Y] - E[Y_{\varepsilon_L}]\|_{L^2(\Omega)} &\leq \|\mathbb{E}[Y] - \mathbb{E}[Y_{\varepsilon_L}]\|_{L^2(\Omega)} \\ &\quad + M_0^{-\frac{1}{2}} \|Y_{\varepsilon_0}\|_{L^2(\Omega)} + \sum_{\ell=1}^L M_\ell^{-\frac{1}{2}} \|Y_{\varepsilon_\ell} - Y_{\varepsilon_{\ell-1}}\|_{L^2(\Omega)}. \end{aligned}$$

3.3 Asymptotic variance convergence rate

We want to bound the variance of the estimator on a level ℓ in the multilevel error bound. In more exact terms, we want to determine for which functions f and domains D the relationship

$$\|Y_{\varepsilon_\ell} - Y_{\varepsilon_{\ell-1}}\|_{L^2(\Omega)} = \mathcal{O}(\varepsilon_\ell^s) \quad \forall \ell > \ell_{\min} \quad (11)$$

holds for some $\ell_{\min} > 0, s > 0$. This behavior is desired since it allows a good multilevel performance as the variance of the estimators on fine levels is already small, thus requiring less fine-level realizations.

Conditioning the expectation on the current position $X_{N_{\ell-1}}$ incorporates the fact that the simulation continues with a given path from level $\ell - 1$, not generating a completely independent one on the finer level ℓ . Using $\bar{X}_{N_\ell} = \bar{X}_{N_{\ell-1}} + \Delta x$ yields

$$\|Y_{\varepsilon_\ell} - Y_{\varepsilon_{\ell-1}}\|_{L^2(\Omega)}^2 = \mathbb{E}[|f(\bar{X}_{N_{\ell-1}} + \Delta x) - f(\bar{X}_{N_{\ell-1}})|^2 | X_{N_{\ell-1}}]. \quad (12)$$

In order to write this expectation in terms of $|\Delta x|$, not in terms of a function of the process we make the following assumption:

Assumption 3.1 (Hölder continuity). *There exist $C, \alpha > 0$ such that*

$$|f(x) - f(y)| \leq C|x - y|^\alpha$$

for all $x, y \in \partial D$.

This assumption implies

$$\mathbb{E}[|f(x + \Delta x) - f(x)|^2 | x] \leq C^2 \mathbb{E}[|\Delta x|^{2\alpha} | x]. \quad (13)$$

The conditioning can be omitted by bounding the right-hand side of (13) by the maximum over x , the stopping point of the process on level $\ell - 1$.

Splitting the expectation into a converging part where $|\Delta x| \leq R_\ell$ and a “diverging” part with $R_\ell < |\Delta x| \leq |D|$, where $|D|$ is the diameter of the circumsphere of D , we get

$$\begin{aligned} \mathbb{E}[|\Delta x|^{2\alpha}] &\leq \mathbb{P}[|\Delta x_\ell| \leq R_\ell] R_\ell^{2\alpha} + \mathbb{P}[|\Delta x_\ell| > R_\ell] |D|^{2\alpha} \\ &= (1 - p_{\text{div}}) R_\ell^{2\alpha} + p_{\text{div}} |D|^{2\alpha}. \end{aligned} \quad (14)$$

Since the $|D|$ in the second term is a domain-dependent constant, we must find a bound for the divergence probability p_{div} in terms of the discretization parameter ε . As shown below, this is possible with the resulting behavior $p_{\text{div}} \propto \frac{\varepsilon_\ell}{R_\ell}$.

In order for the expectation to converge with a certain rate in ε , both terms $R_\ell^{2\alpha}$ and p_{div} should be of the same order. Equating the two, solving

this equation for R_ℓ and inserting back into (14) yields

$$\begin{aligned} R_\ell &= C\varepsilon_\ell^{\frac{1}{2\alpha+1}} \\ \mathbb{E}[|\Delta x|^{2\alpha}] &\leq C\varepsilon_\ell^{\frac{2\alpha}{2\alpha+1}}, \\ \|Y_{\varepsilon_\ell} - Y_{\varepsilon_{\ell-1}}\|_{L^2(\Omega)}^2 &\leq C\varepsilon_\ell^{\frac{2\alpha}{2\alpha+1}}. \end{aligned} \tag{15}$$

For differentiable f , $\alpha = 1$ this yields the convergence rate $\varepsilon_\ell^{\frac{1}{3}}$ in (11).

3.3.1 Convergence of divergence probability

Our goal here is to obtain a bound on the divergence probability p_{div} on a certain level ℓ depending on ε_ℓ . Divergence here means that the process does not result in an estimate $\bar{X}_{N_{\ell+1}}$ that is located within a ball $B_{\bar{X}_{N_\ell}}(R_\ell)$ of radius R_ℓ around the current projected stopping point \bar{X}_{N_ℓ} .

In the derivation we use the so-called harmonic measure. There exist a few equivalent definitions of harmonic measure which are relevant to the current application. The first defines the harmonic measure as a harmonic function satisfying certain boundary conditions, the second is instructive in the context of Monte Carlo approximations of an integral and the third uses a definition relying on the distribution of the first exit point given a Brownian motion. Consider the closed set $D \subset \mathbb{R}^d$ with $d \geq 2$ and let $\mathcal{B}(\partial D)$ denote the σ -algebra of subsets of ∂D and define:

Definition 3.2 (Harmonic measure – Dirichlet solution). *The harmonic measure $\omega_D : D \times \mathcal{B}(\partial D) \rightarrow [0, 1]$, viewed as a function $\omega_D(x, E)$ of x for every fixed D and E , is the unique harmonic function that satisfies the boundary condition [9, p. 39]*

$$f(x) = \begin{cases} 1, & \text{if } x \in E, \\ 0, & \text{if } x \in \partial D \setminus E. \end{cases}$$

Definition 3.3 (Harmonic measure – integral representation). *The harmonic measure $\omega_D : D \times \mathcal{B}(\partial D) \rightarrow [0, 1]$ is the unique function that satisfies [17, sec. 4.3]*

- (a) for each $x \in D$, $E \mapsto \omega_D(x, E)$ is a probability measure on ∂D
- (b) if $f : \partial D \rightarrow \mathbb{R}$ is continuous, the solution of the Laplace equation in D is given by

$$u(x) = \int_{\partial D} f(x') \omega_D(x, dx'). \tag{16}$$

Definition 3.4 (Harmonic measure – hitting probability). *Let X_t^x denotes a Brownian motion started at x . The harmonic measure $\omega_D : D \times \mathcal{B}(\partial D) \rightarrow [0, 1]$ is given by*

$$\omega_D(x, E) := \mathbb{P}[X_\tau^x \in E, \tau < \infty],$$

where $\tau = \inf\{t \geq 0 : X_t^x \in \partial D\}$ is the first exit time of X_t^x from D . [9, Theorem F.6.]

Writing $\omega_D(x, x') = \frac{\partial \omega_D(x, x')}{\partial x'}$ as a probability density function of the distribution of first exit points x' of ∂D by a Brownian motion started at x , (16) corresponds to the expectation value of $f(X_\tau)$. This agrees with the Feynman-Kac formula (2) applied to the Laplace equation. Simulating realizations of the stochastic process X_t thus amounts to approximating the integral (16) by Monte Carlo integration, i.e. generating realizations of the exit points distributed according to the harmonic measure.

Often, the harmonic measure in a certain domain can't be given analytically. Thus, it is important to be able to bound it from above and below. This can be accomplished using the general inequalities popularized by Carleman, illustrated in Figure 1 and often referred to as *Carleman's Principle of Monotonicity* or the *Principle of Extension of Domain*.

Definition 3.5 (Carleman's principle). *The harmonic measure of a portion α of the boundary of $G \subset \mathbb{R}^d$ increases if the region is increased by modifying $\beta = \partial G \setminus \alpha$:*

$$\omega_G(x, \alpha) \leq \omega_{G'}(x, \alpha). \quad (17)$$

Since $\omega_G(x, \alpha) + \omega_G(x, \beta) = 1$, it follows that the harmonic measure on the extended portion of the boundary decreases with respect to the new domain:

$$\omega_G(x, \beta) \geq \omega_{G'}(x, \beta'). \quad (18)$$

[16, p. 68], [7, p. 131].

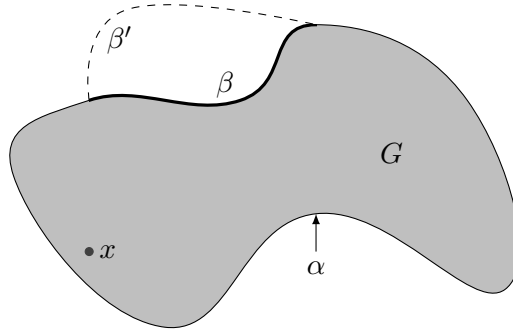


Figure 1: An illustration of Carleman's Principle in two dimensions. α is the unchanged part of the boundary, β is changed to β' . The extended domain is G' , G is the original one.

Writing the divergence probability p_{div} in terms of a harmonic measure $p_{\text{div}} = \mathbb{P}[\bar{X}_{N_{\ell+1}} \notin B_{\bar{X}_{N_{\ell}}}(R_{\ell})] = \omega_D(X_{N_{\ell}}, \partial D \cap B_{\bar{X}_{N_{\ell}}}(R_{\ell})^c)$, we can appeal to *Carleman's Principle*. This allows the transformation of a general bounded convex domain onto a semi-ball, the harmonic measure of which is known analytically. Simple trigonometric observations then result in a bound depending on ε_{ℓ} .

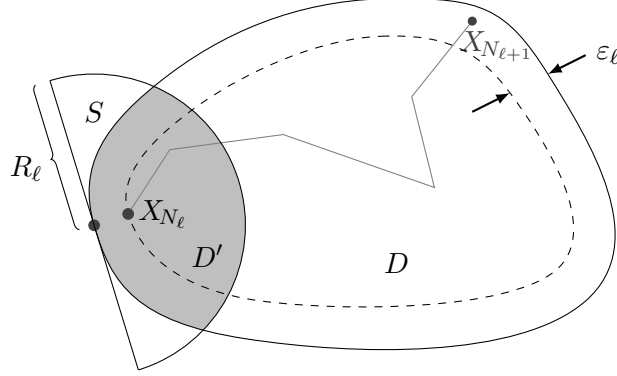


Figure 2: Domain D for simplicity shown in 2 dimensions, with semicircle S of radius R_{ℓ} used to bound the harmonic measure of a point $X_{N_{\ell}}$ located at most ε_{ℓ} from the boundary. The shaded region D' is the intersection of the domain D with the semicircle S . The path to $X_{N_{\ell+1}}$ is “divergent” with respect to $X_{N_{\ell}}$.

The situation in two dimensions is sketched in Figure 2: the intersection of the domain D with a semi-ball S of radius $R_{\ell} > \varepsilon_{\ell}$ is called D' . In Figure 3, the boundaries are split up such that Carleman’s method can be applied to the harmonic measure of $\Gamma_d = \partial D \cap B_{\bar{X}_{N_{\ell}}}(R_{\ell})^c$, the portion of the boundary hit in the case that the process diverges to a point more than R_{ℓ} away from $\bar{X}_{N_{\ell}}$.

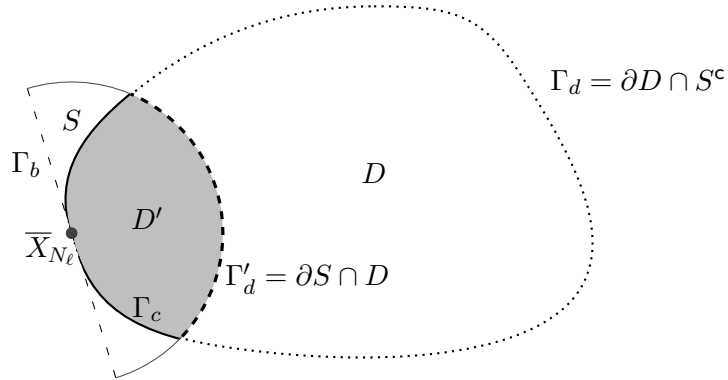


Figure 3: The boundary of the domain D is split into two disjoint portions, the “convergent” boundary Γ_c and the “divergent” boundary Γ_d . The portion of the semicircle boundary ∂S inside of D is denoted by Γ'_d and the base of the semicircle is Γ_b .

We start by bounding the harmonic measure of Γ_d from above by “shrinking” D to D' and using the second bound (18), where $\omega_D(x, \Gamma_d)$ corresponds to the larger domain and thus has smaller harmonic measure:

$$\omega_D(x, \Gamma_d) \leq \omega_{D'}(x, \Gamma'_d).$$

Now, we keep the portion Γ'_d fixed and extend the remaining portion Γ_c of the boundary of D' to the semicircle boundary $\partial D \setminus \Gamma'_d$. This lets us use the first inequality (17), where Γ'_d corresponds to α and Γ_c corresponds to β in the definition, yielding

$$\omega_{D'}(x, \Gamma'_d) \leq \omega_S(x, \Gamma'_d) \leq \omega_S(x, \partial S \setminus \Gamma_b),$$

where the second inequality comes from adding $\omega_S(x, \partial S \setminus (\Gamma'_d \cup \Gamma_b))$, the harmonic measure of the remaining portion of the arc.

We now assume that $R_\ell \gg \varepsilon_\ell$ and that x lies on the normal of the base at its midpoint. Using that the shortest distance between x and Γ_b is less than ε_ℓ allows to use a series approximation of $\omega_S(x, \partial S \setminus \Gamma_b)$ around the midpoint of its base, by using the Dirichlet solution representation of the harmonic measure Definition 3.2. The boundary Γ_b is zero by definition, as illustrated in Figure 4, hence the same is valid for the first and second derivative in all directions of the base. As by definition the Laplacian is zero, the second derivative in normal direction of the base has to be zero as well. Hence the first derivative in this direction is a positive constant. This results in the final bound for the divergence probability in terms of ε_ℓ

$$p_{\text{div}} \leq C \frac{\varepsilon_\ell}{R_\ell},$$

for some constant C .

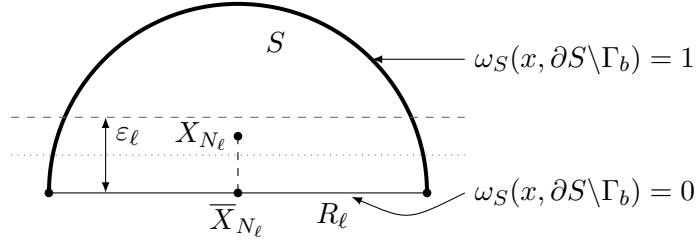


Figure 4: A semicircle containing the point X_{N_ℓ} , for simplicity shown for the 2D case. Γ assumes the value 1 on the arc and 0 on Γ_b .

3.3.2 Optimal number of samples

To simplify notation, we define the estimator Y_ℓ to be

$$Y_\ell := \begin{cases} Y_{\varepsilon_0}, & \text{if } \ell = 0, \\ Y_{\varepsilon_\ell} - Y_{\varepsilon_{\ell-1}}, & \text{if } \ell > 0, \end{cases}$$

and w_ℓ the expected work performed to compute a sample on level ℓ .

Following [15], the optimization problem to solve is then

$$\min \mathbb{E}[W_{\text{tot}}] = \min \sum_{\ell=0}^L M_\ell w_\ell \quad \text{s.t.} \quad \varepsilon_L = \sqrt{\sum_{\ell=0}^L \frac{\text{Var}[Y_\ell]}{M_\ell}}, \quad (19)$$

hence minimizing the expected work, while maintaining a sampling error proportional to the discretization error ($\mathcal{O}(\varepsilon_L)$). [15] provides the optimal number of samples per level

$$M_\ell = \frac{1}{\varepsilon_L^2} \sqrt{\frac{\text{Var}[Y_\ell]}{w_\ell}} \sum_{\ell=0}^L \sqrt{\text{Var}[Y_\ell] w_\ell}. \quad (20)$$

3.4 Error vs. work

The following basic relations will be used in the sequel.

$$\text{Refinement of } \varepsilon_\ell \quad \varepsilon_\ell = \eta^{-\ell} \varepsilon_0 = \eta^{L-\ell} \varepsilon_L \quad (21a)$$

$$\text{Scaling of work} \quad w_\ell = \begin{cases} \mathcal{O}(\varepsilon_\ell^{-\gamma}) \\ \mathcal{O}(\log^p(\varepsilon_\ell^{-1})) = \mathcal{O}(\ell^p) \end{cases} \quad (21b)$$

$$\text{Scaling of variance} \quad \text{Var}[Y_\ell] = \mathcal{O}(\varepsilon_\ell^{2s}) \quad (21c)$$

Determining M_ℓ (20) based on the estimates of variance and work lead to

$$\begin{aligned} M_\ell &= \begin{cases} \varepsilon_L^{-2} \sqrt{\varepsilon_\ell^{2s+\gamma}} \sum_{\ell=0}^L \sqrt{\varepsilon_\ell^{2s-\gamma}} \\ \varepsilon_L^{-2} \sqrt{\varepsilon_\ell^{2s} \cdot \ell^{-p}} \sum_{\ell=0}^L \sqrt{\varepsilon_\ell^{2s} \cdot \ell^p} \end{cases} \\ &= \begin{cases} \varepsilon_L^{-2} \sqrt{\varepsilon_\ell^{2s+\gamma}} \sqrt{\varepsilon_0^{2s-\gamma}} \underbrace{\sum_{\ell=0}^L \sqrt{\eta^{(\gamma-2s)\ell}}}_{\leq C, \forall 2s > \gamma} \\ \varepsilon_L^{-2} \sqrt{\varepsilon_\ell^{2s} \cdot \ell^{-p}} \sqrt{\varepsilon_0^{2s}} \underbrace{\sum_{\ell=0}^L \sqrt{\eta^{-2s\ell} \cdot \ell^p}}_{\leq C} \end{cases} \quad (22) \end{aligned}$$

The summands of the remaining sum can be bounded by a convergent geometric sequence. Thus, by the comparison test the infinite series ($L \rightarrow \infty$) converges to a constant, in the first case provided $2s > \gamma$.

Inserting (22) into the formula $W_L^{\text{tot}} = \sum_{\ell=0}^L M_\ell w_\ell$, we can write the

work as a function of ε_L , which scales linearly with e_{tot}

$$\begin{aligned} W_L^{\text{tot}} &\leq \begin{cases} C\varepsilon_L^{-2} \sqrt{\varepsilon_0^{2s-\gamma}} \sum_{\ell=0}^L \sqrt{\varepsilon_\ell^{2s-\gamma}} \stackrel{(22)}{\leq} C^2 \varepsilon_L^{-2} \varepsilon_0^{2s-\gamma}, & \forall 2s > \gamma \\ C\varepsilon_L^{-2} \sqrt{\varepsilon_0^{2s}} \sum_{\ell=0}^L \sqrt{\varepsilon_\ell^{2s} \cdot \ell^p} \stackrel{(22)}{\leq} C^2 \varepsilon_L^{-2} \varepsilon_0^{2s} \end{cases} \\ &= \mathcal{O}(\varepsilon_L^{-2}). \end{aligned}$$

Thus, we obtain the optimal convergence rate

$$e_{\text{tot}} \propto (W_L^{\text{tot}})^{-\frac{1}{2}}, \quad (23)$$

in the case $w_\ell = \mathcal{O}(\varepsilon_\ell^{-\gamma})$ provided $2s > \gamma$.

3.5 Measured values

The second approach to determining M_ℓ is based on estimating the work w_ℓ and the variance $\text{Var}[Y_\ell]$ on the different levels, rather than using their asymptotic convergence rates as in (22). In the absence of alternatives we still use the asymptotic convergence rate of the discretization error $\mathcal{O}(\varepsilon_L)$. As in [14] the required estimates are computed using the same samples already involved in the MLMC estimator. A certain minimum number of samples (so called warm-up samples) are needed on every level to provide an accurate estimate. Performance disadvantages arise if the number of required warm-up samples exceeds the optimal number of samples (20). This typically happens for the finest level L . However, (22) proves that asymptotically the optimal number of samples grow, which implies that for small ε_L the required warm-up samples do not exceed the optimal number of samples. This technique complicates the implementation slightly, but does not increase the computational work for this application.

4 Numerical results

4.1 Model problems

In this section we formulate model problems, each posed on a different domain to test our algorithm.

4.1.1 Square

Many of the simplest examples are formulated on square domains. Here, we use the two-dimensional square domain $D_\square := [0, 2]^2$. In order to more easily measure the convergence rate, a boundary condition resulting in a

large discretization error is imposed with Hölder exponent $\alpha = 1$.

$$f(x, y) = \begin{cases} 4(x - \frac{1}{2})^2, & \text{if } x \leq \frac{1}{2}, 0 \leq y \leq 2, \\ 4(x - \frac{3}{2})^2, & \text{if } x \geq \frac{3}{2}, 0 \leq y \leq 2, \\ 0, & \text{otherwise, } 0 \leq y \leq 2. \end{cases} \quad (24)$$

The midpoint $m = (1, 1)$ of the square was chosen as the starting point.

4.1.2 Hemisphere

A nice three-dimensional problem to consider is the Laplace equation on a hemisphere $D_\Delta := \{\mathbf{x} \in \mathbb{R}^3 \mid |\mathbf{x}| \leq 1, x_3 \geq 0\}$, as illustrated in Figure 5. The boundary conditions of the Laplace equation are chosen such that the analytical solution is given by $u(\mathbf{x}) = [x_1^2 + x_2^2 + (x_3 + 1)^2]^{-\frac{1}{2}}$, as stated in [13]. This leads to the following formulation with Hölder exponent $\alpha = 1$:

$$\begin{aligned} \Delta u &= 0 \text{ in } D_\Delta \\ u(\mathbf{x}) &= [2(x_3 + 1)]^{-\frac{1}{2}} \text{ for } |\mathbf{x}| = 1 \\ u(\mathbf{x}) &= [x_1^2 + x_2^2 + 1]^{-\frac{1}{2}} \text{ for } x_3 = 0. \end{aligned} \quad (25)$$

The initial point was chosen to be $x_0 := (0.2, 0.3, 0.1)$.

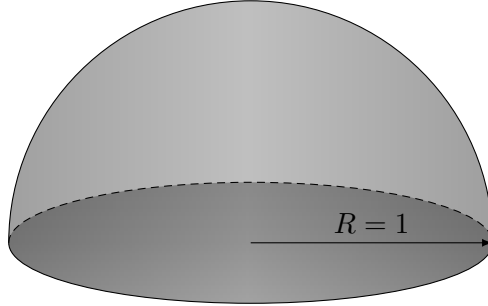


Figure 5: Hemisphere domain D_Δ in three dimensions.

4.2 Implementation

A generic C++ implementation for both the conventional and the multilevel version of the Walk on Spheres algorithm was created. The M_ℓ are either determined analytically or based on measurements during the simulation. The code uses MPI to generate samples in parallel, such that larger problems can be solved. The code was tested with up to 128 cores. Further information on the implementation is found in [8].

4.3 Measurement methodology

Monte Carlo methods are based on the approximation of an expectation by a sample mean. The resulting estimator is itself a random variable – thus, the total error given by

$$\|\mathbb{E}[Y] - E[Y_{\varepsilon_L}]\|_{L^2(\Omega)} = \mathbb{E}[\|\mathbb{E}[Y] - E[Y_{\varepsilon_L}]\|^2]^{\frac{1}{2}}.$$

This outer expectation can again be approximated by a sample mean, corresponding to a repeated call of the corresponding algorithm. Thus, for a certain set of parameters one must call the implementation with different random number seeds and compute the sample mean over the resulting realizations. An estimate for a confidence interval can also be computed using the results of these repeated calls.

4.4 Measured variance convergence rate

As stated in (11), the variance of the estimate on a level ℓ is assumed to converge like $\mathcal{O}(\varepsilon_\ell^s)$. In (15), it was shown that s is bounded by $\frac{\alpha}{2\alpha+1}$, α being the exponent of Hölder continuity of the boundary condition. The results presented here are empirical measurements showing the convergence of the variance of the estimator. The measurements were averaged over ten calls for the square domain D_\square . In Figure 6 the theoretical and the measured exponent s are shown. Shown only for the square domain, but

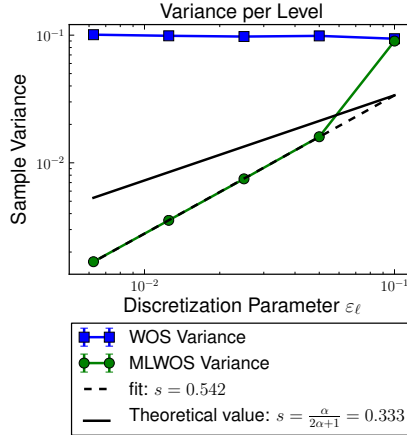


Figure 6: D_\square Convergence of the variance as the discretization parameter ε_ℓ is reduced for the square domain D_\square with $\alpha = 1$ using $\eta = 2$.

valid for all described domains, the variance convergence rate is around 0.5, which is better than the value $\frac{1}{3}$ obtained from the analytical derivation assuming continuity of the boundary condition. This is not a contradiction since the derivation yields an upper bound. However, using the analytical

upper bound will very likely result in suboptimal results, since obviously one can get away with fewer samples on the fine levels.

4.5 Measured convergence rate

In this section we compare the error versus work of the proposed multilevel WoS algorithm with the plain WoS algorithm. For the multilevel WoS algorithm we distinguish between two strategies, one with analytically derived M_ℓ , w_ℓ and $\text{Var}[Y_\ell]$ (MLWOS) and the other with M_ℓ based on measured w_ℓ and $\text{Var}[Y_\ell]$ (MEAS) according to Section 3.5.

Measurements performed on the domain D_\square are shown in Figure 7. We performed measurements with $\eta = 2$, $\eta = 8$ and $\eta = 16$. The measured values for the average error and the average work are shown, together with the 1σ confidence interval, for different algorithms. Sampling was performed until the confidence interval was very small. In all cases the multilevel WoS with analytically derived M_ℓ (MLWOS) performs poorly compared to the plain WoS algorithm. The MLWOS performance improves for $\eta = 16$, but an improvement over the plain WoS algorithm is not measured. The multilevel WoS with M_ℓ using measured values (MEAS) always performs better than the plain WoS algorithm, especially for $\eta = 16$ where the computation is up to 2 times faster.

The analytically derived M_ℓ are based on the $\text{Var}[Y_\ell]$. As observed in Section 4.4 $\text{Var}[Y_\ell]$ converges faster than predicted in our theory. Hence it is not surprising that the multilevel WoS algorithm with analytically derived M_ℓ performs suboptimally. The multilevel WoS algorithm with M_ℓ based on measurements does not suffer from this problem, therefore we observe a better performance.

The parameter η does not influence the convergence rate of the multilevel scheme, but a clever choice may asymptotically reduce the work by a constant. Therefore it is expected that the performance of the multilevel WoS algorithm depends on η . For certain problems (see e.g. [10]), this parameter is optimized analytically. Here, various values are tried on the domain D_\square in order to empirically find good values. We observed that $\eta = 16$ is a better choice than the other tested $\eta = 8$ and $\eta = 2$.

Figure 8 shows measurements for the domain D_\triangle for $\eta = 16$. The results are similar to the ones seen for the domain D_\square . The multilevel WoS with analytically derived M_ℓ (MLWOS) performs poorly, where as the multilevel WoS with M_ℓ using measured values (MEAS) is up to 2 times better than the plain WoS algorithm.

5 Conclusions and outlook

In this work, a first application of the Multilevel Monte Carlo method to the Feynman-Kac formula was developed. A novel version of the Walk on

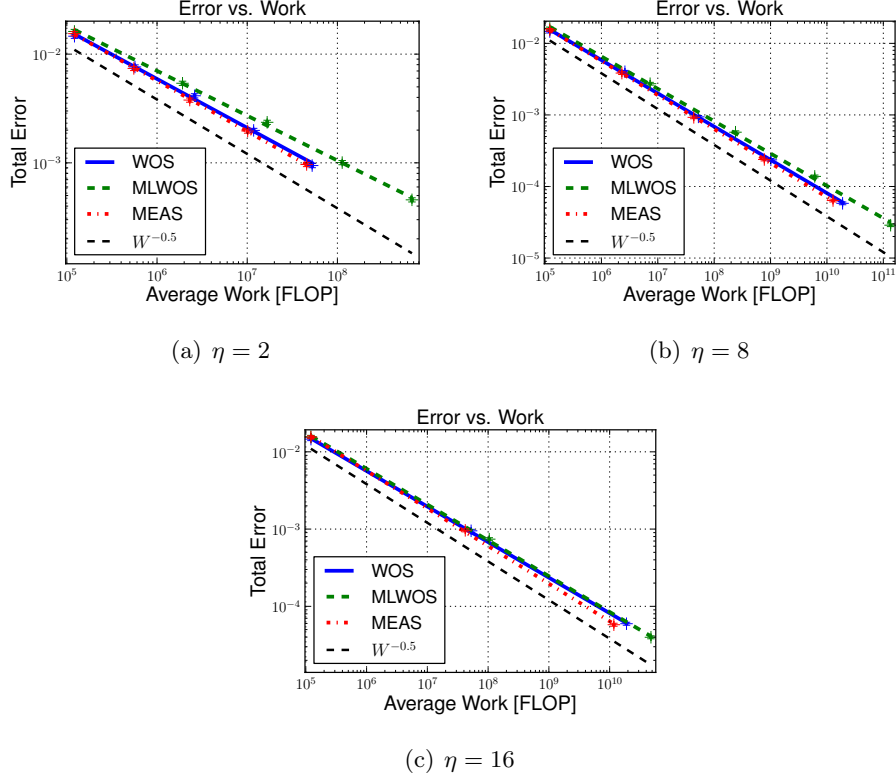


Figure 7: Convergence of error vs. work for the problem posed on D_{\square} , for various values of the refinement parameter η .

Spheres process, the “Multilevel Walk on Spheres” was formulated and central quantities were derived. It was proven that the “Multilevel Walk on Spheres” algorithm convergence rate of error versus work is optimal.

In order to measure the actual performance gain of the proposed method, an implementation of both the conventional and the multilevel version of the Walk on Spheres algorithm was created. By parallelizing the generation of samples, more thorough convergence results could be obtained. Additionally, a version of the multilevel method that chooses the number of samples based on measurements was implemented.

In order to test a variety of different situations, multiple domains and boundary conditions were implemented. The convergence rate of the error with respect to the work was measured, along with many other relevant quantities. “Multilevel Walk on Spheres” resulted in a reduction of the work by up to a factor 2.

Using tighter bounds, it would perhaps be possible to find an upper bound for the variance convergence rate that better fits the measured behavior. This would hopefully allow to analytically determine the number of

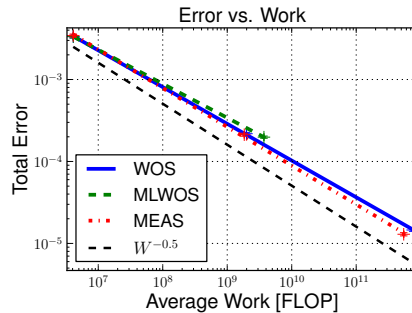


Figure 8: Convergence of the error vs. work for the problem posed on D_{Δ} , using $\eta = 16$.

samples per level.

It would be beneficial to generalize the “Multilevel Walk on Spheres” to other elliptic equations, for example the Poisson equation with nonconstant right-hand side. This equation has applications in many fields, such as particle accelerator modeling or biochemical electrostatics, in which a Multilevel Monte Carlo formulation may be of use. This would require the formulation of a process that contains multiple discretization levels *along the path*, i.e., with level dependent maximal sphere radius. This would increase the expected work especially for fine discretizations and presumably increase the benefit of MLMC over MC.

References

- [1] A. Adelmann, P. Arbenz, and Y. Ineichen. A fast parallel poisson solver on irregular domains applied to beam dynamics simulations. *Journal of Computational Physics*, 229(12):4554–4566, 2010.
- [2] A. Barth, C. Schwab, and N. Zollinger. Multi-level Monte Carlo Finite Element method for elliptic PDEs with stochastic coefficients. *Numerische Mathematik*, 119(1):123–161, 2011.
- [3] I. Binder and M. Braverman. The complexity of simulating Brownian Motion. *Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 58–67, 2009.
- [4] N. Bruijn. *Asymptotic Methods in Analysis*. Bibliotheca mathematica. Dover Publications, 1970.
- [5] F. M. Buchmann and W. P. Petersen. Solving Dirichlet problems numerically using the Feynman–Kac representation. *BIT Numerical Mathematics*, 2003.

- [6] J. Chang. Stochastic Processes <http://www.stat.yale.edu/~jtc5/251/stochastic-processes.pdf>. *Lecture notes*, 2007.
- [7] J. L. Doob. *Classical Potential Theory and Its Probabilistic Counterpart*. Classics in Mathematics Series. Springer, 2001.
- [8] R. Gantner. Computing the Feynman-Kac formula efficiently with multilevel Monte Carlo. *Master thesis, Computer Science Department, ETH Zurich*, March 2013.
- [9] J. B. Garnett and D. E. Marshall. *Harmonic measure*. Cambridge University Press, 2005.
- [10] M. B. Giles. Multilevel Monte Carlo Path Simulation. *Operations Research*, 56(3):607–617, 2008.
- [11] N. Hilber, O. Reichmann, C. Schwab, and C. Winter. *Computational Methods for Quantitative Finance: Finite Element Methods for Derivative Pricing*. Springer finance. Springer, 2013.
- [12] M. Kac. On distributions of certain Wiener functionals. *Transactions of the American Mathematical Society*, 65(1):1–13, 1949.
- [13] M. Mascagni and C.-O. Hwang. ε -Shell error analysis for “Walk On Spheres” algorithms. *Mathematics and Computers in Simulation*, 63(2):93–104, June 2003.
- [14] M. E. Muller. Some continuous Monte Carlo methods for the Dirichlet problem. *Annals of Mathematical Statistics*, 27(3), 1955.
- [15] F. Müller, P. Jenny, and D. W. Meyer. Multilevel Monte Carlo for two phase flow and Buckley–Leverett transport in random heterogeneous porous media. *Journal of Computational Physics*, 250:685 – 702, 2013.
- [16] R. H. Nevanlinna. *Analytic functions*. Princeton mathematical series. Princeton University Press, 1960.
- [17] T. Ransford. *Potential Theory in the Complex Plane*. Cambridge University Press, 1995.
- [18] J. G. Wendel. Hitting spheres with Brownian motion. *The Annals of Probability*, 8(1):164–169, 1980.